# Co-Designing QuickPic: Automated Topic-Specific Communication Boards from Photographs for AAC-Based Language Instruction

Mauricio Fontana de Vargas
mauricio.fontanadevargas@mail.mcgill.ca
McGill University
Montréal, Canada

Christina Yu
christina.yu@childrens.harvard.edu
Boston Children's Hospital
Boston, USA

Howard C. Shane
howard.shane@childrens.harvard.edu
Boston Children's Hospital
Boston, USA

Karyn Moffatt
karyn.moffatt@mcgill.ca
McGill University
Montréal, Canada

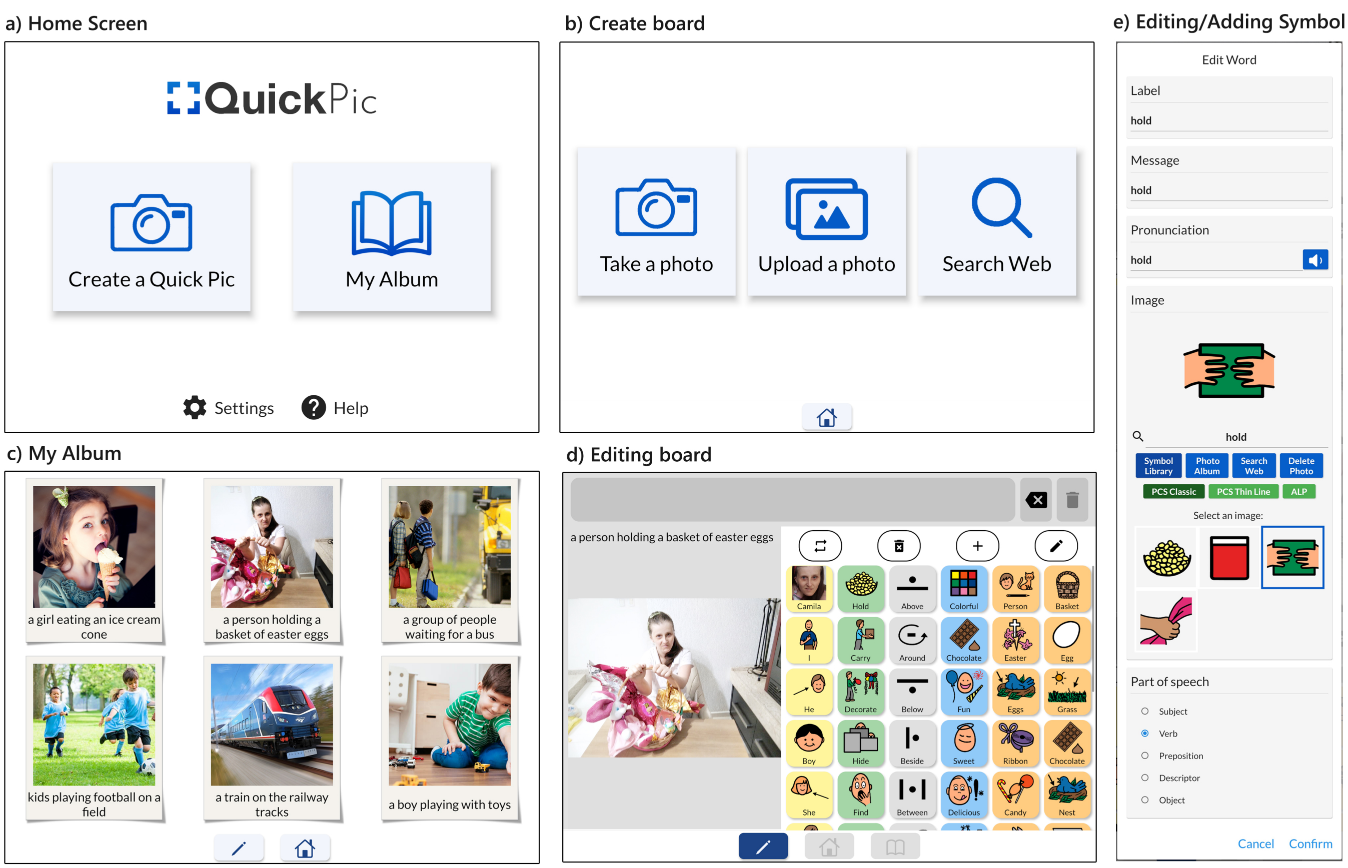


**Figure 1: QuickPic generates topic-specific communication boards automatically from photographs to enable speech language professionals and special education teachers delivering "just-in-time" language support to non-speaking individuals, alleviating the demands of creating personalized language support material. Users can also edit the board generated automatically in a streamlined fashion.**

## ABSTRACT

Traditional topic-specific communication boards for Augmentative and Alternative Communication (AAC) require manual programming of relevant symbolic vocabulary, which is time-consuming and often impractical even for experienced Speech-Language Pathologists (SLPs). While recent research has demonstrated the potential to automatically generate these boards from photographs using artificial intelligence, there has been no exploration on how to design such tools to support the specific needs of AAC-based language instruction. This paper introduces QuickPic, a mobile AAC application co-designed with SLPs and special educators, aimed at enhancing language learning for non-speaking individuals, such as autistic children. Through a 17-month design process, we uncover the unique design features required to provide timely language support in therapy and special education contexts. We present emerging evidence on the overall satisfaction of SLPs using QuickPic, and on the advantages of large language model-based generation compared to the existing technique for automated vocabulary from photographs for AAC.



## CCS CONCEPTS

• **Human-centered computing** → **Accessibility systems and tools**; **Empirical studies in accessibility**; **Human computer interaction (HCI)**.

## KEYWORDS

Augmentative and Alternative Communication, autism, just-in-time, LLM, assistive technology


## 1 INTRODUCTION

Speech-Language Pathologists (SLPs) and special education teachers play a pivotal role in enhancing the expressive communication abilities of autistic children and other individuals with complex communication needs, often using Augmentative and Alternative Communication (AAC) as a tool within therapy sessions and classrooms. A typical approach relies on providing communication boards with vocabulary symbols relevant to a specific topic grouped in a single display. These boards are then used by the professionals in activities that immerse learners in symbolic language, for example, by demonstrating how to select symbols to talk about real-world concepts and stimulate learners to formulate sentences for requesting, answering, and commenting during interesting and meaningful moments [44]. A primary challenge for these professionals is the work associated with manually selecting vocabulary and programming it into the AAC device in preparation for these activities. This process is extremely time-consuming and often infeasible [6, 13]. Even with extensive experience and accumulated practice, speech-language professionals often allocate a significant portion of their workload to programming AAC tools, which reduces the time available for interacting with clients [32, 52]. Similarly, family members report the amount of effort and time required to program the devices to do similar activities at home as a major factor contributing to the abandonment of AAC systems [4, 47].

Recent work by Fontana de Vargas et al. [15] has demonstrated the potential of applying artificial intelligence to automatically generate these communication boards, presenting a substantial opportunity to decrease the time and complexity involved and the ability to provide "just-in-time" (JIT) communication support [40]. While their work indicated an interest in using JIT generated boards in AAC-based therapy, their methodology did not permit a deep investigation into the professionals' design requirements for user interface or vocabulary generation quality to properly provide language support for their clients. Instead of following a participatory design approach, researchers designed a prototype (Click AAC) to serve as a probe for a broader investigation into how AAC tools with vocabulary generated from photographs could be utilized in natural contexts, powered by a generation method originally designed for autobiographical storytelling [11].

Therefore, it is unknown how AAC tools offering topic-specific boards automatically generated from photographs should be designed to meet the needs of professionals who rely heavily on providing immersive visual supports for non- and minimally-speaking individuals. It is also unclear to what extent the language generation method originally adopted in Click AAC is able to provide vocabulary appropriate for the therapy context, given that researchers only performed a simulation of communication performance based on a storytelling dataset [11] and a qualitative investigation into the broad impressions of vocabulary quality [15]. Furthermore, the rapidly changing landscape in the artificial intelligence field since the introduction of the method raises the question of whether the automation on topic-specific boards can benefit from the generative capabilities of large language models (LLMs).

In this work, we address these open questions by following a research through design approach [58]. We co-designed QuickPic, a mobile AAC application for assisting SLPs and special education professionals creating topic-specific communication boards to support children learning and using symbolic language during therapy and school activities. Over a 17-month period, our design team formed mainly by a HCI researcher, four SLPs, and two special education teachers participated in an interactive process to elicit required features, design the application interactive interface, and investigate the relevance of vocabulary generated automatically from photographs by the method originally used in Click AAC [11, 15] and a novel method based on the LLM GPT 3.5. Following on this, we conducted a user study with another 8 SLPs to compare their overall experience when creating boards under the two generation methods.

This work contributes:

(1) QuickPic, the first AAC tool co-designed with experienced SLPs and special education teachers able to generate topic-specific communication boards automatically from photographs. This innovation reveals unique design features necessary to effectively assist professionals in delivering "just-in-time" language support to their learners.
(2) Emerging evidence on the superior appropriateness of GPT-based generation of vocabulary for the context of AAC-based language instruction, in comparison to the method originally employed in Click AAC [11, 15].

# 2 RELATED WORK

## 2.1 AAC-based Speech Language Therapy

Augmentative and Alternative Communication (AAC) tools can enhance communication for nonspeaking individuals, thereby offering improved social interaction and independence. While individuals primarily facing physical barriers to communication (e.g., adults with Amyotrophic Lateral Sclerosis, ALS) can compose complex and nuanced sentences through text-based systems, those with Autism Spectrum Disorder (ASD), other developmental disabilities, or lexical processing impairments such as aphasia must usually rely on less expressive symbol-based systems where vocabulary is represented by symbols or images [26, 27].

Speech-language therapy focused on the use of AAC tools plays an important role in enabling these individuals to learn how to communicate through symbolic language. This involves instructing the association between real-world objects or concepts and their symbolic representation, as well as the syntactic structures (e.g., subject-verb-object) necessary to formulate meaningful messages from a combination of individual symbols. One of the main activities performed toward these goals is the *aided language stimulation*[1], in which professionals compose sentences themselves in the AAC tool while speaking. Topic-specific communication boards, also known as topic-displays [17], are a valuable AAC tool to facilitate emerging communicators engaging in those activities. Because a small set (e.g., 1–30) of vocabulary related to a specific topic or activity is displayed in a single page, often grouped by part of speech, learners (and professionals performing aided language stimulation) can access vocabulary and compose messages without having to navigate to other pages, reducing overall cognitive and memory demands when communicating [19, 25, 52].

A well-recognized clinical approach that prominently incorporates topic-displays and other forms of symbolic AAC is the Visual Immersion System (VIS) [13, 41, 43, 44]. This method was formulated in response to research findings and clinical observations indicating that autistic individuals exhibit relatively robust aptitude in visual processing [2] and manifest a pronounced disposition for engaging with visual content delivered via electronic screens [42]. The VIS approach, as well as vast clinical evidence [25–27], emphasizes the necessity of immersing learners within an environment rich in visually symbolic elements across various settings, including home, school, and community, paired with ample opportunities for practical utilization—similarly to the immersive setting used for sign language communication in the deaf community [1] and for individuals learning a second spoken or written language [10]. These visual elements are then used for various communication purposes, such as expressing preferences, making requests, giving directions, making comments, asking questions, and understanding social cues. To adhere to these principles, it is essential to use tools capable of quickly and conveniently generating and presenting visual content, such as topic-displays, on a nearly constant basis [37]. This kind of instantaneous support, often referred to as "just-in-time" (JIT), facilitates learning by reducing the demands on working memory and leveraging situated cognition during teachable moments [40], thus facilitating seamless day-to-day communicative exchanges. However, to date, very little exploration has been undertaken on techniques for automating, and consequently, fully conveying JIT support.

[1]also known as aided language input or modelling.

## 2.2 Improving AAC Tools: Efforts from the HCI Community

*2.2.1 Text-based AAC.* Research aimed at enhancing text-based AAC systems focus primarily on applying Natural Language Processing (NLP) techniques to expedite message composition. Given the motor and cognitive demands associate with text-entry for people with physical disabilities, communication rates of these AAC users tend to be around 1–25 words per minute, while spoken communication typically happens at a rate of 150–200 words per minute [9].

One common approach, often employed in commercial AAC tools such as the Tobii Dynavox Communicator[2], relies on language models to anticipate the next character or word users intend to input based on their previous message context [22, 49, 50], akin to predictive text on smartphones. User's contextual information has demonstrated to be valuable for enhancing these prediction algorithms. By using users personal information, such as the nature of their disability, age, literacy level, or user's environment data, such as current location, time, or conversation partner's messages [21, 45, 53], researchers have demonstrated key-stroke rate gains up to 71%. Contextual information was also explored for the generation of sentences to help children with disabilities telling stories to their parents about school activities[38, 48].

Researchers have also explored how to expand compressed, telegraphic user input into coherent sentences, going from initial efforts from the 1990's using basic NLP techniques such as semantic parsing [12, 30], to recent research taking advantage of LLMs [8, 51]. Notably, most participants in the Valencia et al. [51]'s work found that a prompt-based generation of sentences from single words to be "very or extremely useful", potentially helping them to create requests for routine, self-care and accessibility related tasks in a daily basis.

*2.2.2 Symbolic AAC.* Exploration on techniques aimed at improving AAC support for individuals with developmental disabilities, such as Autism Spectrum Disorder (ASD) and cerebral palsy, as well as for those with lexical impairments caused by stroke (i.e., aphasia) and who therefore require symbolic representation of language concepts, is less extensive. Research in the field primarily focuses on techniques aimed at empowering users to efficiently access a large number of symbols they require for diverse communication needs and various contexts. This could potentially alleviate the burden on speech-language professionals and family members, who currently have to manually select and program relevant vocabulary into the devices. Shin et al. [46], for exmple, designed a system to support caregivers in sharing information about AAC use by children with complex communication needs in order to better support them, encouraging balanced participation between caregivers and reducing cognitive loads.

[2]https://www.tobiidynavox.com/pages/communicator-5-ap

Artificial intelligence and context-aware computing provide opportunities to address the vocabulary access issue through automated provision of symbolic vocabulary. Early investigations in this realm involved Wizard of Oz experiments on context-aware applications that retrieves previously programmed vocabulary based on the user's current location, conversation topic, or conversation partner [20]. More recently, Obiorah et al. [36] introduced an application that translates photos of food and menus into interactive symbols using image captioning and optical character recognition (OCR) to assist individuals with aphasia when dining in restaurants. However, their technique cannot generate symbols for concepts not directly depicted in the photographs. In an effort to create a broader range of vocabularies from photographs to aid individuals with aphasia in retelling past activities, Mooney et al. [31] proposed a method that involves extracting common words from human comments on social networks where the user's photographs are posted. Nevertheless, their study only tested a proof-of-concept system using a simulated social network.

Advancing the approach of automatic generation of vocabulary from photographs to support autobiographical storytelling, de Vargas and Moffatt [11] proposed the use of the Visual Storytelling Dataset (VIST) [18] as the main source of vocabulary. VIST is composed of 65,394 photos of personal events, grouped in 16,168 stories. Each photo is annotated with captions and narrative phrases that are part of a story, created by Amazon Mechanical Turk workers. Their method operates by initially identifying the photographs in VIST that are most similar to the input photograph. This is achieved by calculating sentence similarity between the input photo caption, generated using the computer vision technique from Microsoft Azure (Computer Vision API v 3.1[14]), and all VIST photo captions. Subsequently, the method retrieves all stories associated with those photographs and determines the most relevant words to suggest to users.

This technique was later employed in the Click AAC [15] prototype, enabling researchers to uncover the dynamics between individuals with complex communication needs, their conversation partners, and automated language support across the contexts of speech language therapy and special education. However, it is important to note that the Click AAC design did not involve the participation of end-users (i.e., SLPs). The app was created as a probe to broadly investigate how AAC tools with vocabulary generated from photographs could be used in those natural contexts. Consequently, it remains unclear to what extent the autobiographical storytelling method proposed by de Vargas and Moffatt [11] is capable of providing suitable vocabulary and whether its interactive language support meets the needs of speech-language professionals in the creation of materials to support AAC-based language therapy.

## 3 PHASE 1: CO-DESIGN OF QUICKPIC

### 3.1 Method

To elicit the unique design features and requirements necessary for effectively assisting professionals in delivering "just-in-time" language support to their learners, we adopted the research through design approach [58], where the construction of tangible artifacts and systems is emphasized as the central means for knowledge creation [3]. Consequently, the resulting designs reflect designers' judgments on how to address the inherent possibilities and challenges in their field [16].

To embark on the system construction, we formed a design team composed of professionals with HCI, speech-language therapy, and special education backgrounds, and followed the methodological framework by Scaife et al. [39]. This framework delineates four key phases in the co-design process: a) identifying the issues present in current practices, discussing conventional technologies used within the domain, and establishing the fundamental requirements for the new technology; b) translating these requirements into software specifications and assessing their feasibility; c) creating low-tech prototypes, and finally, d) designing and testing high-tech prototypes to assess interactivity and verify whether the developed prototypes offer enhancements over existing technologies.

We conducted this iterative process over a 17-month period, as illustrated in Figure 2. Throughout this timeframe, we conducted internal testing among the design team members, and collected insights from professionals after informal testing of the application with individuals with communication disabilities. We iterated on the design three times until all parties involved were satisfied with the application. We now introduce the design team, followed by a detailed description of each part of the design process.

### 3.2 Participants

Two main groups took part in the design process, as described below. The profiles of the individual team members are presented in Table 1. In addition to them, during the last iteration round, a group of seven special education teachers was also invited to test QuickPic and provide their opinions regarding the app usability.

(1) **Initial design team:** one HCI researcher experienced developing assistive technologies, and two SLPs (SLP-1, SLP-2) who work directly, in a daily basis, with non- and minimally-speaking children who use or are learning to use symbolic communication. They are the first three authors in this paper.
(2) **Invited special education (SPED) professionals:** Two SLPs (SLP-3, SLP-4) and two special education teachers (TCH-1, TCH-2) joined the team in the second round of iteration to provide their opinions and contribute to the app design. These professionals were invited by the two SLPs from the initial team to test the prototypes during their regular school activities with their autistic students. This testing process uncovered necessary usability improvements and required features for the successful adoption of the technology in the school setting. SLP-1 and SLP-2 have maintained an ongoing relationship with the invited professionals for more than seven years. They have been collaborating in the delivery of cutting-edge instructional methodology based on visual strategies to teach students with autism and other communication disorders. They regularly meet once a month to discuss these strategies and novel technologies that have the potential to support their students.

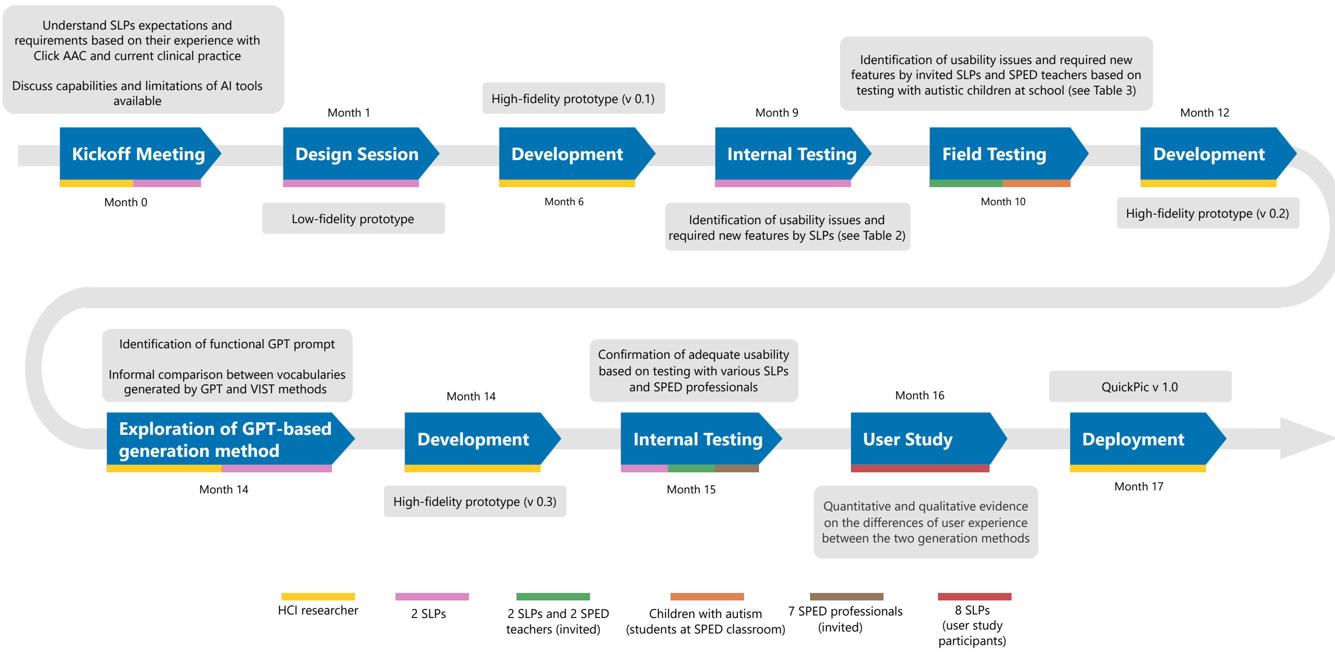


**Figure 2: The design process of QuickPic. At each step, the design team met online to discuss ideas, share progress, and define next steps. Month numbers indicate when a certain stage was completed. Colored bars indicate the parties involved in a step.**

## 3.3 The co-design process

*3.3.1* ***Kick-off Meeting.*** The initial design team met via video conferencing to enable the HCI researcher to comprehend the SLPs' initial requirements for general features and their expectations regarding the characteristics of automatically generated vocabulary. SLP-1 and SLP-2 highlighted that a common barrier for them and other SLPs was the substantial amount of time required in creating topic specific displays on top of their current caseloads. They also explained the importance of constructing boards around topics of personal interest to their clients, which captures their attention during activities and consequently leads to better learning outcomes.

To ensure that the SLPs' expectations aligned with what could realistically be developed within the project timeframe, the design team deliberated on the capabilities and limitations of the AI technologies available at the time. This included computer vision techniques available for image captioning and object identification (e.g., Microsoft Azure Computer Vision, Google Cloud Vision, AWS Rekognition), and natural language generation methods for AAC. Based on this discussion and the SLPs' previous experience with Click AAC [15], the team decided to use an updated version of the image captioning model used in Click AAC (i.e., Microsoft Azure Computer Vision 3.2) as a starting point, along with the same vocabulary generation method (hereafter referred to VIST [11]), since it was the only method specifically designed for and used in the AAC context at that time. This method uses as input the caption provided the captioning model.

*3.3.2* ***Low-fidelity Prototype Design.*** Following the insights gathered during the kick-off meeting discussion, the two SLPs engaged in an iterative process to design a low-fidelity prototype that encapsulates all the necessary screens and interface interactions to effectively convey the envisioned language support. This prototype, as illustrated in Figure 3, comprises four main screens. The rationale for their designs is given below:

**a) Create my board:** SLPs found it valuable to offer users the ability to create new boards by either importing existing photos from the device's storage, taking new photos with the device's camera, or searching for images online.

**b) Main display:** Once the input photograph is processed, the application should display the resulting communication board. The SLPs' design closely resembles the layouts of existing topic-display AAC apps. It features a message bar at the top with the symbols of the message being composed, the central photograph contextualizing the board's main topic, and a flat grid containing vocabulary symbols categorized by part of speech (subjects, verbs, prepositions, descriptors, and objects). The SLPs determined that the symbols presented should be primarily from a popular AAC symbol set, widely employed in commercial AAC tools: the Picture Communication Symbols (PCS) library. Furthermore, they stressed the importance of allowing users to edit the generated communication board. They should have the freedom to remove, reorder, edit, or add symbols, accomplished through the "edit mode" button found in the bottom menu.

**c) Edit or Add a Symbol:** Once in "edit mode," users should be able to specify a series of parameters for a symbol, including: i) the symbol label (the word displayed under the symbol), ii) the message

| ID | Relevant Experience |
|---|---|
| HCIR | PhD with 7 years of experience conducting research on and developing assistive technologies for individuals with communication disabilities. Has volunteered in a local support group for people with aphasia, assisting speech-language pathologists in delivering communication activities to the group members. Has a sibling with disabilities who communicates through unaided AAC (e.g., gestures, pointing). |
| SLP-1 | Msc, Licensed Speech Language Pathologist and holder of the ASHA Certificate of Clinical Competence (CCC), with 7 years of experience delivering language therapy for AAC users and candidates. Current case load of 250 pediatric patients per year. Clinical coordinator of the Autism Language Program at Boston Children's Hospital, one of the largest programs for autistic children in the world. Conducts research on AAC for autism and has published several papers in top-tier venues on developmental disabilities and AAC. |
| SLP-2 | PhD with over 30 years of experience in delivering AAC-based language therapy and creating assistive technologies for children and adults with communication disabilities. Has successfully managed and directed large research and clinical groups dedicated to AAC practice. Is the founder of the Augmentative Communication Program and the Autism Language Program at Boston Children's Hospital, an institution that has developed more than a dozen computer applications and has established a clinical program for supporting autistic children using technology. Conducts research on visual support for autistic persons and has an extensive list of publications in the field. |
| SLP-3 | CCC-SLP with 25 years of dedicated practice in the field, specializing in daily AAC utilization. Proficient in crafting topic-specific displays, typically implemented daily at the start of the school year, with ongoing weekly updates throughout the academic calendar. |
| SLP-4 | CCC-SLP with 22 years of experience working with AAC users in a daily basis. Creates topic-specific communication boards weekly. |
| TCH-1 | Special education teacher and Board Certified Behavior Analyst (BCBA). Has more than ten years of experience providing home base therapy and behavioral therapy with autistic children and young adults. |
| TCH-2 | Special education teacher with more than ten years of experience. |

**Table 1: Profiles of the design team members. The first three members are authors in this paper and actively participated in the design process since the beginning. The other four members were brought to the team after the development of the first high-fidelity prototype.**

(the string added to the message bar, which may be lengthy), and iii) pronunciation (the message reproduced through synthesized speech, often used to correct the pronunciation of people's names). Furthermore, users must have the option to select an image from the device's storage or the internet in case the PCS symbol is not deemed adequate. Finally, users must be able to choose in which column the symbol should appear based on the word's part of speech.

**d) QuickPic Library:** The SLPs determined that users should be able to view and access all boards previously created through a dedicated screen. On this screen, each board is represented by a polaroid-style photo containing the board title, generated automatically.

In addition to these main screens, the SLPs recognized the need for personalization of the interface and vocabulary complexity to cater to different users' abilities. As a result, they designed a settings menu with options for selecting: i) the number of items generated for each part of speech, ii) the background color of each part-of-speech column, iii) voice type, rate, and pitch, and iv) vocabulary generation level, which includes three choices: elementary (only symbols describing the photograph), intermediate (a small set of words closely related to the photograph), and advanced (a larger set of words related to the scene). According to them, this distinction allows them to avoid overwhelming children who are just starting to learn symbolic communication through activities where the goal is to formulate sentences describing what is happening in the photographs, while also enabling more proficient children to expand their language using vocabulary that is not directly depicted in the photograph but it is still related to the topic.

#### 3.3.3 *Internal Tests with QuickPic v. 0.1.*

Once the HCI researcher had developed a high-fidelity prototype that replicated the SLPs' initial design, SLP-1 and SLP-2 tested the application by themselves to investigate whether the envisioned features were correctly implemented and to explore the quality of vocabulary generated.

This was primarily achieved by inputting photographs commonly used in their therapy activities, such as "a dog running on a beach" or "a girl eating ice-cream," and reviewing the vocabulary symbols generated to ensure their relevance to the topic. They compiled a list of necessary changes in a shared spreadsheet with illustrative images to clearly communicate their requirements to the HCI researcher. Through email exchanges and brief video calls, the team collectively discussed the new features and necessary usability improvements, presented in Table 2.

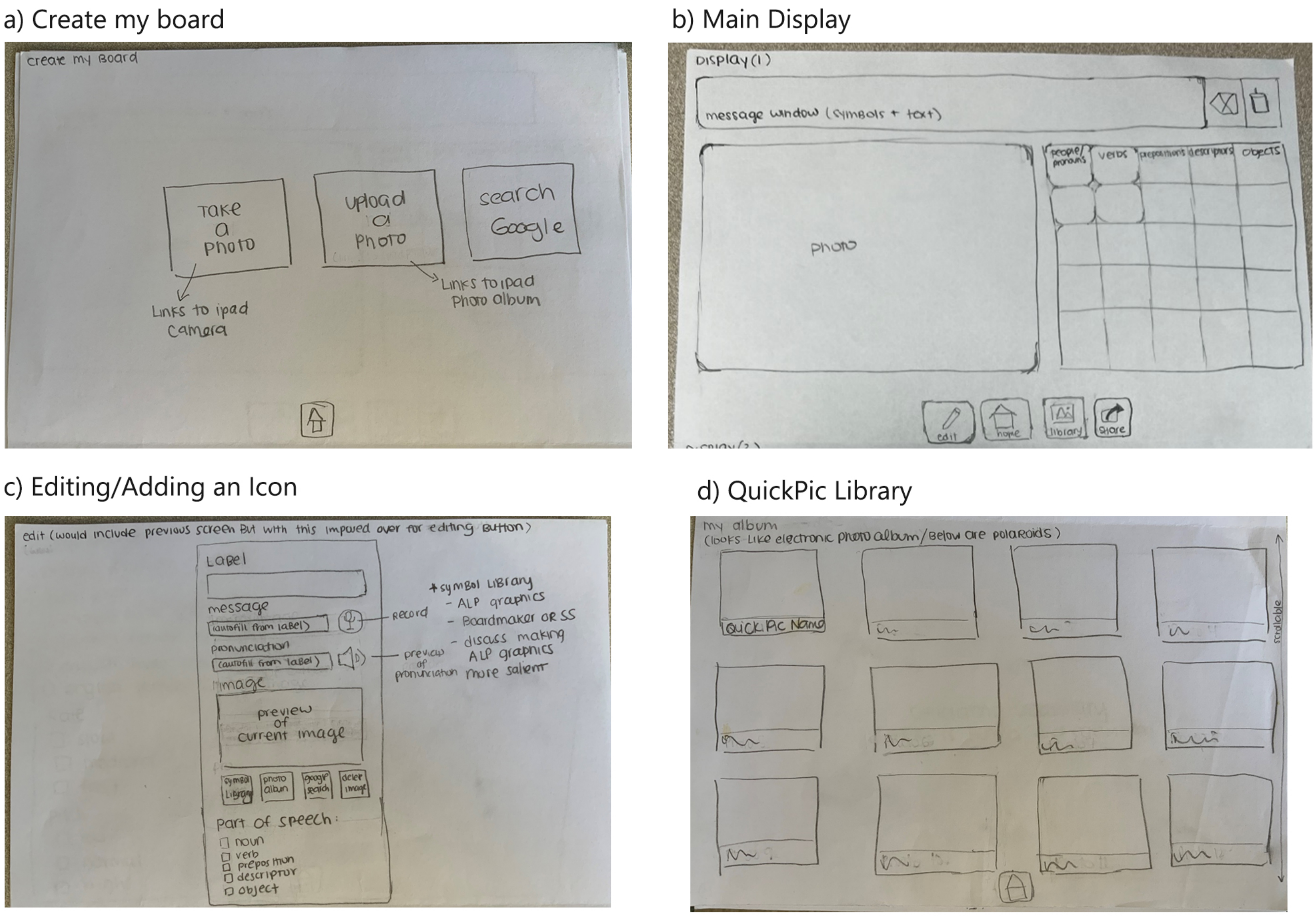


**Figure 3: QuickPic's low-fidelity prototype designed by SLP-1 and SLP-2.**

| Type | Description |
|---|---|
| Generation | **1.** Ensure all words in the automatically-generated photo caption are included in the symbols grid.<br>**2.** Automatically include the names of familiar people in the subjects column.<br>**3.** Improve relevance of generated descriptors: frequently, the generated descriptors lack a strong connection to the input photograph, with recurring descriptors such as "little," "old," and "young" appearing across various photos.<br>**4.** Enhance part-of-speech classification accuracy: too many words are added in the wrong column. |
| Interface | **5.** Incorporate ALP animated symbols to effectively convey descriptors, prepositions, and verbs, as autistic individuals may struggle with comprehending static visual representations of abstract concepts.<br>**6.** Allow users to select the default symbol set (i.e., PCS or ALP) used when creating new boards.<br>**7.** Automatically apply a symbol when editing a vocabulary item based on the entered word (label) and the default symbol set.<br>**8.** Ensure that the symbol used to represent automatically generated words aligns with the word's part-of-speech. For example, 'paint' should be represented as the action of painting if it is categorized under the verbs column and as the object 'paint' if it is placed under the objects column. |

**Table 2: New features and necessary usability improvements identified by SLP-1 and SLP-2 after testing QuickPic v. 0.1.**

#### 3.3.4 *Field Tests with QuickPic v. 0.1 by Invited SPED Professionals Involving Autistic Children*.

Following the internal tests conducted by SLP-1 and SLP-2, the four invited SPED professionals (SLP-3, SLP-4, TCH-1, TCH-2) tried QuickPic for a few days during their routine activities at school. We did not provide specific instructions on how or when to use the app but rather let the professionals decide, as opportunities for language support naturally arose. They used QuickPic to provide customized, just-in-time language support to their non- and minimally-speaking students, primarily autistic children. More specifically, professionals created boards *on the fly* using the method (i.e., taking a new photo, uploading from the device's library or searching the web) they found better suited for the moment. They then used the boards to model symbolic language while talking about a certain topic with the students (i.e., aided language stimulation).

This field test did not intend to collect data from app usage; instead, its purpose was to assess initial impressions, overall usability, and the experience in the special education classroom context. Through online meetings and email exchanges, the team discussed overall experiences and identified commonly mentioned additional features and usability issues. According to the invited professionals, QuickPic needed to be easier and quicker to use than what was commercially available and what they were already familiar with, in order for it to be successfully used in the classroom. While they all agreed that their experience with the app was positive, they uncovered seven important requirements that had not been identified during the previous internal testing. Table 3 outlines the features, usability concerns, and necessary improvements raised by these SPED professionals.

#### 3.3.5 *Technical Development of the Second Iteration (QuickPic v. 0.2)*.

In addition to the interface-related improvements, some points regarding vocabulary generation were successfully implemented using simple approaches and publicly available tools. Feature number 2 was implemented through a dedicated screen accessed under the app settings. Users can input sample photographs of familiar people and indicate their names. When generating new boards, the photograph is then analyzed by a facial identification service[3] to attempt matching with the previously configured individuals. If a positive match occurs, their names are added to the "subjects" column. For item number 11, the team established a straightforward rule-based mechanism that incorporates pronouns based on keywords found in the photo captions. For instance, if the caption contains words like "boy" or "man," the pronouns "I," "he", and "boy" are added. If words such as "people", "children", or "family" appear in the caption, the pronouns "I," "he", "she", "we", "they", "boy", and "girl" are included. If no person is in the photo, only the pronoun "it" is added. Finally, item 12 was addressed by integrating into the generation mechanism a list of all 787 common adjectives and adverbs obtained from the Corpus of Contemporary American English (COCA)[4] dataset, paired with their opposites obtained through ChatGPT. Therefore, if the user enables the "opposite" feature in the settings menu, for each descriptor word generated, its opposite is also added to the board.

However, the group identified that addressing most of the high-priority issues related to the vocabulary generation (i.e., items 3, 4, 8, 9, 10) would necessitate an entirely new generation method. Furthermore, item 8 could not be implemented due to the absence of part of speech information in the PCS and ALP libraries.

#### 3.3.6 *Exploration of a New Vocabulary Generation Method*.

Motivated by the recent public release of the ChatGPT platform—which occurred two months prior to the testing conducted by the SPED professionals—we embarked on an exploration of an alternative method for generating vocabulary for QuickPic leveraging the capabilities of `gpt-3.5-turbo` by Open AI, the state-of-the art LLM available through an API at the time.

In fact, SLP-1 and SLP-2 were already recommending in their clinical practice the utilization of ChatGPT to SLPs, parents, and teachers as a valuable tool for generating vocabulary ideas for children's communication devices. To address the common issue of finding appropriate verbs and descriptors when customizing topic-specific displays, they utilized straightforward prompts like "provide me with 10 verbs related to bus", or "list 10 basic adjectives for playground". As a result of discussions with the SLPs, the HCI researcher proceeded to refine the prompt design based on recommendations from the language model community and drew upon their own experience working with LLMs on other projects. This refinement led to the creation of the following prompt:

> You are a Speech Language Pathologist specialized in Augmentative and Alternative Communication.
> Your task is to provide vocabulary related to a situation to help a person with communication disability to formulate messages about the situation. This vocabulary must contain words that people would often use to talk about that situation, either to describe it or to discuss it in a general context.
> The vocabulary must contain 20 verbs, 20 descriptors (adjectives and adverbs not terminating with LY), 20 objects, and 20 prepositions.
> All words must be in the first person singular, infinitive form without 'to'.
> Provide your answer as a JSON object as in the following example:
>
> EXAMPLE ANSWER:
> "verbs": [VERBS],
> "prepositions": [PREPOSITIONS]
> "descriptors": [ADJECTIVES AND ADVERBS NOT TERMINATING WITH LY],
> "objects": [OBJECTS],
>
> SITUATION: [*Photo caption from the computer vision model is inserted here*]

The team conducted a subjective comparison between the vocabulary generated using the novel method, and the original one by de Vargas and Moffatt [11], hereafter referred to as GPT and VIST, respectively. This comparison was based on input from 13 different photo descriptions (i.e., situations) that had been frequently used by the design team in previous testing sessions[5].

[3]Microsoft's Azure Cognitive Services.
[4]https://www.english-corpora.org/coca/
[5]"kids playing soccer", "a table full of food", " a group of people standing in front of a school bus", "a girl eating ice-cream", "a boy playing on the sand", "a person riding a

| Type | Description |
|---|---|
| Generation | **9.** Reduce the presence of irrelevant words in the vocabulary generated under the "advanced" setting.<br>**10.** Enhance the "intermediate" and "advanced" generation to provide more relevant words for common school contexts, such as "children sitting in front of a school bus".<br>**11.** Generate pronouns based on photo content rather than relying on fixed pronouns [I, you, he, she, we, they].<br>**12.** Provide users with the option to enable the generation of opposite descriptors (e.g., hot–cold) in the settings. |
| Interface | **13.** Allow users to toggle the display of generated captions on or off, as it can be either helpful or distracting depending on the student's abilities.<br>**14.** Provide users with the option to display the generation settings used for each board if they wish to view them.<br>**15.** Give users the option to show or hide the photograph when viewing the communication board because some children may get distracted, or overwhelmed by the photograph when interacting with symbols. |

**Table 3: New features and necessary usability improvements identified by invited SLPs and SPED professionals after using QuickPic v. 0.1 with their students at special education classrooms.**

*3.3.7* ***QuickPic v. 0.3 and Final Validation Testing.*** Given that SLP-1 and SLP-2 found that the GPT generation method appeared to provide more relevant vocabulary compared to the original VIST method, the group decided to integrate GPT into the app. Users were given the option to select their preferred method for creating vocabulary boards through the app settings. This modification marked the development of the third version of the high-fidelity prototype (QuickPic v. 0.3), depicted in Fig. 1.

By the fifteenth month of the project, to reduce biases that the parties actively involved in the design process might have had and to broaden the set of test cases in real-world contexts, we distributed the QuickPic v. 0.3 to an additional seven special educators within the same school district as the invited SPED professionals in the design team. These professionals, along with the entire design team, tested the app by themselves using photographs representing topics commonly discussed in the classroom context. The objective was to assess the overall usability and vocabulary quality of QuickPic when utilizing the GPT method, and whether QuickPic would speed up the creation of personalized boards in comparison to existing AAC tools they already use. At this point, all parties involved were confident that QuickPic could be effectively employed in language and school contexts without the need for significant new features or usability improvements, as commented by the invited SPED teachers: "vocabulary generation is appropriate and accurate", "very intuitive to program and use", "great way to create materials on the fly".

## 4 PHASE 2- USER STUDY

While our co-design process revealed professionals' apparent satisfaction with the application usability and the superiority of the GPT method over VIST [11], it remained unclear to what extent these impressions were unbiased, given the extensive involvement of the project stakeholders. Therefore, we proceeded to the second part of our research agenda, where we conducted a user study with a new group of SLPs who had no involvement in the design process.

The study goal was to gather evidence regarding the overall usability and user experience of QuickPic, in addition to the relevance of vocabulary produced, when powered by the two generation methods. To accomplish this, we provided a prompt describing a typical scenario encountered in their clinical practices and asked the SLPs to create communication boards using QuickPic under the two conditions. This study was approved by the Boston Children's Hospital Institutional Review Board (IRB). We now present the study details.

### 4.1 Participants

We recruited eight SLPs who work with children with developmental disabilities at the Center for Communication Enhancement (CCE) at Boston Children's Hospital based upon convenience sampling. To qualify for inclusion, the participants needed to meet the following criteria: i) possess an active ASHA Certificate of Clinical Competence for Speech-Language Pathologists (CCC-SLP), ii) have a minimum of one year of experience working with individuals who use, or are candidates for AAC, iii) have prior experience in creating topic-specific displays for their therapy activities. None of the recruited SLPs were involved in co-design and preliminary tests presented in Section 3. Table 4 presents an overview of the participants' profiles.

### 4.2 Procedures

**a) Pre-questionnaire:** participants completed a questionnaire regarding their demographic information and past AAC experience. They also responded to a 2-point Likert scale questionnaire concerning the benefits of topic-specific boards and the challenges associated with their creation using current AAC applications.

**b) Tutorial:** participants received a printed QuickPic's reference guide (developed by SLP-1 and SLP-2) and were instructed to go through it independently to become familiar with QuickPic's features. Following this, they engaged in a one-to-one quick tutorial

chicken", "a Christmas tree", "a cow and a dolphin jumping on the sea", "a dog running on a beach", "a group of people waiting the bus", "a person sitting on a bench", "a person skiing", "a skier going down a slope".

| ID | Age category | Years of SLP practice | Frequency working with AAC users | Frequency creating topic displays | Average time needed to create a topic display (min) |
|---|---|---|---|---|---|
| P1 | 25-34 | 2 | Weekly | Occasionally | 31-40 |
| P2 | 25-34 | 4 | Daily | Weekly | 11-20 |
| P3 | 35-44 | 17 | Daily | Monthly | < 10 |
| P4 | 25-34 | 12 | Daily | Weekly | 11-20 |
| P5 | 25-34 | 2 | Weekly | Monthly | 21-30 |
| P6 | 34-44 | 12 | Daily | Daily | < 10 |
| P7 | 25-34 | 6 | Daily | Monthly | 11-20 |
| P8 | 55-64 | 35 | Monthly | Occasionally | 51-60 |

**Table 4: Demographic information of participants in our user study.**

session in which SLP-1 demonstrated the features mentioned in the reference guide, such as creating a new board, editing a board, and modifying individual buttons.

**c) Board creation:** participants engaged in the creation of communication boards using QuickPic. Using the "search web" function, each participant created one board under the GPT condition and another under VIST, both centered around the topics "car" or "train", as chosen by the participant. The two boards created did not necessarily need to be exactly the same, but were about the same topic. The following prompt was provided to participants to guide them in this task:

> You have a seven year old male patient with a primary diagnosis of autism spectrum disorder - level 3. Medical history includes no functional concerns regarding vision, hearing, or motor. Receptive language skills include strong comprehension of noun-based vocabulary and ability to follow single-step directions within familiar contexts. Expressive language skills include scripted phrases (e.g., I want *something* ), and single word approximations to label. Aided communication strategies include a grid-based communication application used primarily for requesting, labeling, and protesting. A goal of speech therapy is commenting and describing using three-word utterances. A highly preferred activity/topic of conversation are cars and trains. Based upon this scenario, create a topic-specific display revolved around cars or trains using QuickPic's "search" function.

While the exact goals of therapy will vary according to the child's abilities, this prompt was designed by the SLPs in the team to provide a typical scenario encountered in their clinical practices. The car/train topic is of high-interest for autistic patients, and are frequently accessed during therapy. For both conditions, QuickPic was configured to generate a maximum of 10 items per part of speech, and with the opposite generation of descriptors turned on. Participants were not informed about the specific generation method being used (i.e., GPT or VIST), and the order of conditions was randomized among participants to minimize any potential effects related to the order. The initially generated and final vocabulary boards (after participants' editing) were then saved as images for the posterior analysis of the vocabulary considered appropriate by participants.

**d) Post-questionnaires:** after using QuickPic, participants filled out two questionnaires comparing the usability and overall experience under the two conditions. The first questionnaire was the Mobile-health App Usability Questionnaire (MAUQ [57]), a psychometrically validated instrument to capture users perception of the usefulness and usability of health applications, including those for therapeutic and education purposes[6]. The second was a 7-point Likert scale post-questionnaire adapted from Fontana de Vargas et al. [15] to assess the user experience in terms of interaction, vocabulary quality, and overall usage.

**e) Open-ended questions** : participants responded to five questions via email, providing feedback on their overall experience with QuickPic, opinions about the two conditions, thoughts on incorporating the tool into their clinical practice, and any usability improvements or new features they would like to have in future versions of the app.

## 4.3 Findings

*4.3.1 Attitudes Towards Topic-Specific Displays.* All eight participants agreed with the following benefits of topic-specific boards: i) facilitates expansion of utterance length, ii) supports clinicians in addressing communication goals during sessions, iii) assists clinicians in modeling symbolic vocabulary, and iv) enhances clients' ability to communicate about specific topics. Additionally, six participants (75%) believed that topic-specific boards improved the fluidity of communication about a specific topic. Regarding the obstacles to the boards use, all participants noted time constraints as a major barrier to including topic-specific displays in their sessions. Other barriers showed varying levels of consensus: three participants (37.5%) found it challenging to create visually appealing boards and were uncertain about the organization, framework, and guidelines for their use; two (25%) had difficulty identifying vocabulary and suitable language, and one reported lacking the necessary resources (e.g., apps, software) to create the boards.

[6]Four questions out of the original 21 were removed due to their inapplicability to the AAC domain.

*4.3.2 Automatic Captioning Quality.* We first analyze the captions automatically generated from the photographs chosen by participants when creating the boards to better understand whether the computer vision model used in QuickPic was able to produce accurate descriptions. We found that the QuickPic's captioning model was able to correctly identify the scene in all the five different photos chosen by the eight participants.

To gain a better understanding of whether the model update influenced the quality of vocabulary presented to users, we also assessed the predecessor model from Click AAC using the same input photographs. Our analysis revealed that the automatic captioning model from Click AAC failed to accurately describe the central elements in two photographs. It generated the caption "a child playing with blocks" instead of "a child playing with a toy train" and "two boys sitting on a toy tractor" instead of "two boys sitting on toy trains." The remaining three photos were correctly captioned by both computer vision models, yielding the descriptions "a blue sports car driving on a road," "a boy playing with toys," and "a couple of boys playing with toy cars".

*4.3.3 Comparison of Vocabulary Generated by GPT and VIST.* Relevant vocabulary was defined as the number of vocabulary symbols generated automatically that each participant kept on a finalized board. Six participants kept more symbols under the GPT condition, while the remaining two (P6, P7) kept the same number across conditions. Overall, for a given participant, the finalized boards consisted mostly of the same vocabulary across conditions. To provide a more detailed view of the quality of vocabulary generated, Fig. 4 shows the mean number of symbols maintained and the mean number of symbols manually added by the participants for each part of speech. While a statistical test of significance is not adequate due to the number of data points, this analysis provides emerging evidence of the superiority of the GPT method across most part of speech categories. The GPT method generated more relevant words than VIST for all parts of speech, with a substantially larger difference observed in prepositions (4.1 vs 1.4), followed by objects (3.0 vs 1.5), descriptors (3.6 vs 2.3), and finally verbs (3.1 vs 2.8). For subjects, which were generated using the same algorithm on both two conditions, the number of items kept was similar (2.3 vs 2.1).

The small difference for verbs can be explained by the nature of VIST generation. Since the method outputs frequent words from stories about personally relevant events, generic verbs such as "go", "get", and "see" (also known as core vocabulary in the AAC realm) tend to be generated independently of the input photograph. Consequently, they are also useful for the photographs used in the therapy context and were retained by the SLPs in the final boards. In contrast, objects and descriptors tend to be more specific to a given situation. The substantial difference in preserved prepositions can be explained by the low number of prepositions generated by VIST across all the photographs in the study. On average, only 2.5 prepositions were generated per photograph, while GPT tended to generate a number close to the maximum allowed (10).

The GPT method also outperformed VIST when considering the proportion of symbols generated automatically in relation to the entire vocabulary present in the final boards. Using GPT, participants manually added fewer items than the number of relevant symbols generated across all parts of speech. In contrast, when

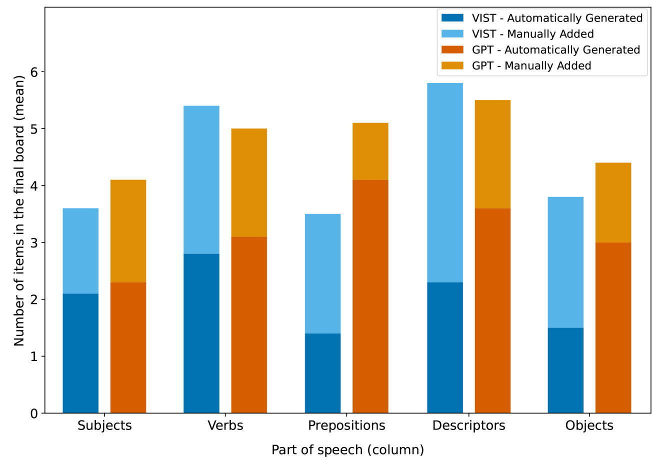


**Figure 4: Comparison of the number of symbols kept on the final boards between GPT and VIST conditions.**

using VIST, participants manually included more symbols than the number of relevant symbols generated for descriptors, prepositions, and objects.

*4.3.4 Comparison of Usability and Overall Experiences.* In general, participants created the boards faster under the GPT condition, taking from 3 min 36 s to 9 min 56 s ($\bar{x}$ = 6 min 13 s). When using VIST generation, participants took from 4 min 4 s to 21 min 53 s ($\bar{x}$= 9 min 39 s).

Fig. 5 presents the post-questionnaire responses in the form of a diverging stacked bar chart, with participant answer counts on the x-axis. The horizontal bars are centered around the neutral category (4). Overall, participants expressed satisfaction with the experience and support provided by QuickPic. However, there was a preference for the experience offered by GPT-based generation. Major differences were observed in responses related to the quality of generated vocabulary. Under the VIST condition, most participants did not feel that generated vocabulary included words they wanted to use nor that the vocabulary was effective in helping them achieving targeted goals, while they tended to report the opposite for GPT (questions 6 and 10). Likewise, a majority of participants did not believe that the vocabulary generated by VIST included relevant words they would not have thought of by themselves, while most participants reported that GPT generation helped them to expand the vocabulary present in the final boards effectively (question 8). Interestingly, even when using VIST , most participants agreed that they could create topic-displays more efficiently with QuickPic compared to traditional AAC tools.

Fig. 6 presents the results from the MAUQ questionnaire. Overall, usability scores ranged from 2.4 to 7.0, ($\bar{x}$ = 4.8) for the VIST condition, while scores ranged from 4.1 to 7.0, ($\bar{x}$ = 5.5) on the GPT condition, suggesting overall higher usability on the GPT condition. While the creators of the MAUQ questionnaire [57] have not specified a minimum threshold for determining good usability, previous research [33] suggests that average scores below 4.0 indicate poor usability. Consequently, our findings clearly demonstrate

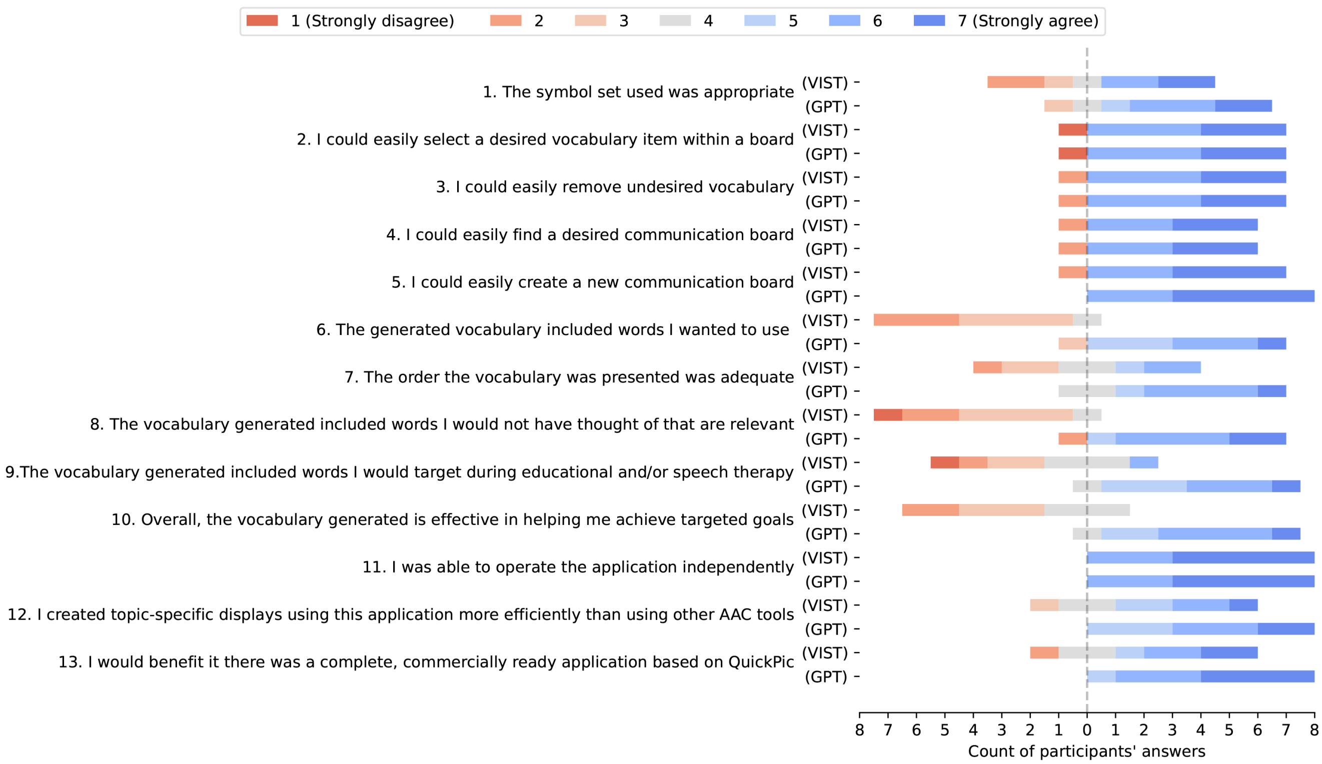


**Figure 5: Post-questionnaire scores comparing the experiences under VIST and GPT generation methods.**

that QuickPic, particularly when powered by GPT generation, has strong usability.

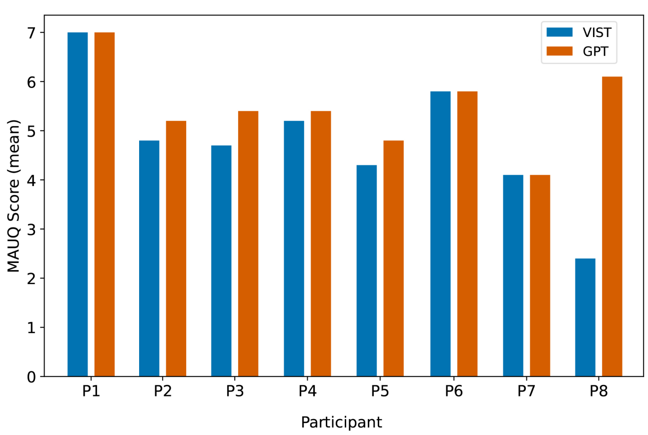


**Figure 6: MAUQ mean scores comparing the overall usability under GPT and VIST generation methods.**

*4.3.5 Open Ended Questions.* The responses gathered from participants uniformly underscore the efficacy of QuickPic for expediting and simplifying the creation of topic-specific communication boards, mainly because "the app provided a starting point" (P1), and "editing was simple and effective" (P4).

Remarkably, three participants stated that their experience with QuickPic surpassed that of standard AAC apps, citing that "it was easier to make topic display boards than using Boardmaker or TouchChat HD-AAC" (P1), "it was much easier and quicker to program in comparison to another AAC app I have used" (P5), "Boardmaker is going to go out of business" (P8).

All eight participants expressed their willingness to incorporate QuickPic into their practice. They identified several potential use cases, such as "creating displays on the fly for common activities" (P4) for "patients' interests" (P2), "in a much more efficient manner ... help[ing] me increase aided language modeling in sessions" (P5). Participants also mentioned the usefulness of QuikPic to "help families independently select vocabulary at home" (P7). Lastly, desirable features suggested by participants included the customization of the symbols' skin tone and the ability to save commonly used words for quicker access when editing the boards.

# 5 DISCUSSION

## 5.1 Unique Features and Requirements Uncovered

Our design process revealed specific requirements not present in the Click AAC app nor in the findings from Fontana de Vargas et al.'s study [15], both regarding the quality of vocabulary generated and the interactive interface, shedding light on the strengths of co-designing with experienced end users.

In past research, Click AAC was introduced "as a generic tool aimed at supporting a wide set of contexts" [15] to broadly investigate how symbolic vocabulary generated automatically from photographs could support learners and professionals in natural contexts of speech language therapy and special education. In addition to insights regarding the overall concept, their study revealed that vocabulary generated by their specific implementation of the vocabulary generation algorithm (named VIST in this paper) was able to provide desired words, as reported by 7 out of 9 participants who used the English version of Click AAC. However, this significantly departs from our findings. Even with an updated computer vision model capable of correctly identifying all photographs in our user study (including two that were not captioned correctly by the model used in Click AAC), 7 out of the 8 participants in our study reported that vocabulary generated by VIST did not have words they wanted to use.

This difference highlights the importance of our co-design approach. The VIST method can be useful for supporting autobiographical storytelling of personally-relevant events [11]—which can be extremely helpful for adults with aphasia [23, 31]—or providing imaginative vocabulary that can be used as a springboard to expand language or stimulate conversation through other forms of AAC [15]. However, our co-design process revealed that this method is not well-suited to provide words closely related to topics commonly discussed during therapy and special education.

The explanation for why VIST generation did not provide proper language support for the context of this study relies on the lexicon dataset used as to extract vocabulary related to a photographed scene. Their technique was built on the Visual Storytelling Dataset [18] (hence the name VIST), which is formed by sentences for telling stories about events photographed and posted on a social network (i.e., Flickr), such as a family dinner or a friend's birthday party. This leads to two issues in the context of QuickPic usage: **i)** the lack of photos similar to the ones used in therapy and special education (e.g., "kids sitting in front of school bus", "a child playing with a toy train"), meaning that their algorithm will not pull up similar photographs from the VIST dataset, and **ii)** for those scenes with similar photos in the dataset, the associated lexicon used to extract vocabulary from has a storytelling nature, containing imaginative words that are loosely related to the scene photographed, while SLPs require words more closely related to the photographs.

In addition to the aforementioned vocabulary quality aspects and new features (e.g., automation of new types of vocabulary such as subjects, familiar people, and opposite descriptors), our design process revealed requirements related to the interactive interface that can substantially support professionals in creating topic-specific boards in a timely manner. Indeed, even under the VIST generation condition, most professionals in our user study felt that they created the boards more efficiently using QuickPic in comparison to current AAC technologies they often use. This is due to the interface designed by our team, as it enables users to quickly import photographs from the internet to represent the central topic of the board and add new words while associating them with symbols from different libraries or the internet in a streamlined fashion.

## 5.2 Reflections on the Co-Design Process

The highly positive usability and user experience expressed by the SLPs during the final rounds of internal testing and in our user study confirmed that our co-design process resulted in a refined prototype, ready to meet the professionals' needs. We can attribute this success not only to the vast experience of the SLP team members in delivering AAC-based therapy to individuals with complex communication needs, but also to the origin of our collaboration.

A remarkable fact is that our collaboration stemmed from the initial interest of the SLPs and their home institution, which led to a strong commitment and enthusiasm from the professionals throughout the technology development. They actively supported and endorsed the project, successfully bringing additional groups of professionals from other contexts to validate their ideas, provide additional insights, and test the prototypes. The inclusion of new professionals across different stages of the project was possible due to SLP-1 and SLP-2's experience managing large groups of SLPs who work in the same context, and training special education teachers from partner institutions. These cycles of idea generation, paired with validation by other professionals, guaranteed that a wide range of views were captured throughout the design process, which consequently led to the unveiling of a range of novel features.

It is also important to note that the insights brought to the team by the HCI researcher were not limited to technical aspects. His experience, gained from previous interactions with other SLPs and people with aphasia from local support groups, as well as his personal experience communicating with family members through AAC, and his background in developing similar AAC apps and interacting with SLPs from several countries during the development and research of those apps, provided the means for effective communication with the rest of the design team. It also gave him a proper understanding of the problem they were addressing and the ability to suggest relevant avenues for exploring the usability of the application during both the design process and the final user study.

Since our goal was to design an AAC tool primarily for use by SLPs and SPED teachers to provide language support and instruction, rather than for use by individuals with disabilities themselves, we decided to include only professionals (the end-users) in the design team. People with communication disabilities, such as autism, participated indirectly through field tests. While this approach was fruitful, it is important to note that the active participation of people with disabilities in the design decisions could have led to more comprehensive findings and a different set of AAC tools. To overcome communication challenges that may arise during the design process with autistic individuals, future research could explore approaches aimed at empowering users through an accessible design space, relying on non-verbal artifacts and interactions, as successfully employed when co-designing not only with autistic people [54, 55] but also with people with aphasia [34, 35, 56], dementia [28], and older adults [29]. Naturally, the differences in communication abilities among these populations raise the question of the extent to which and how different approaches can be adapted to foster effective and creative design spaces.

## 5.3 Opportunities for Technical Improvements

This work has contributed significant advancements compared to previous research and commercially available AAC tools. Nevertheless, the rapid evolution of artificial intelligence and context-aware computing offers immediate prospects for further enhancing the support provided by automated AAC applications.

*5.3.1 Improving the Automatic Association of Symbols.* A critical requirement identified during our design process was the accurate assignment of vocabulary symbols based on the part of speech of a given word. Some English words, such as "work" and "paint" can function as both verbs and nouns, necessitating distinct symbols for each part of speech. Existing symbol libraries, like the Boardmaker PCS used in this work, lack part of speech labels among their 40,000 symbols. This lack of information has not been an issue so far for traditional AAC devices because they require users to manually select symbols when adding a new word. However, this limitation poses additional challenges for automated AAC systems, necessitating non-trivial solutions.

A potential avenue for addressing this challenge in future research involves leveraging machine learning models in conjunction with crowdsourcing. Initially, a computer vision model can analyze all images in the symbol library, generating descriptive tags and sentences for each symbol. For instance, the word "paint" might result in descriptions like "a bucket full of paint" for the noun representation and "a person standing in front of an easel" for the verb representation. Subsequently, a NLP algorithm, potentially powered by LLMs, can categorize these descriptions into different part-of-speech categories and associate them with the corresponding input image. Since the application of machine learning techniques, as proposed, can be prone to errors, a second step could involve crowd-sourced workers verifying and correcting the initial classification.

*5.3.2 Improving Identification of Photographs Content.* An issue that past research on the topic [15] has revealed, but has not been fully explored in this work, is the incorrect or incomplete captioning of photographs often employed in therapy and school contexts. Our analysis of captions generated for a small set of photographs chosen by professionals in our user study, combined with the fact that none of the professionals emphasized the need for better image identification during the design process, suggest that using the Microsoft Azure Vision v. 3.2 model likely reduced misidentifications compared to previous solutions like Click AAC. However, since the development of QuickPic, novel computer vision models (e.g., BLIP2[24]) have been released, yielding state-of-the-art performances. Subsequent work should investigate whether these models can provide more accurate descriptions for improved identification of image content in the AAC context.

*5.3.3 Improving Vocabulary Generation.* The recent release of several LLMs, including GPT 4[7], PaLM 2[8], and LLaMA 2[9], opens up opportunities for enhancing vocabulary generation in AAC tools. Future research should investigate whether integrating these models into symbol-based AAC systems can lead to more contextually relevant and coherent vocabulary suggestions. Another avenue for exploration is the customization of prompts "on the fly" through the app settings to include the learner's profile, encompassing their interests, sensitive topics, and preferred vocabulary. This personalized approach to vocabulary generation could significantly enhance the relevance and coherence of suggestions tailored to each user's unique needs and preferences.

## 5.4 Limitations and Future work

This work primarily focused on the design process of QuickPic and the study of its usability, complemented by the first quantitative investigation of the quality of topic-displays generated automatically. While our methodology provided valuable insights into usability and user experiences under two different methods, there remain several limitations and avenues for future research to explore for further advancement in symbol-based AAC tools.

*5.4.1 Comprehensive Evaluation of Vocabulary Generation Methods.* Future work should conduct a more comprehensive evaluation of vocabulary generation methods for symbolic AAC. This could potentially unveil new features and improvements that may not have been uncovered by our study participants. An important avenue to explore is an extensive investigation of photographs taken by SLPs and special education teachers during real activities with their learners. Publicly releasing applications like QuickPic provides an opportunity to collect large datasets of photographs used to generate communication boards, along with the resulting boards after the users editing. These data could then be turned into publicly available datasets, facilitating the assessment of future generation methods without the need for extensive user studies. Computer simulations could be used to assess vocabulary relevance using part of these datasets as ground-truth, expediting the development and selection of novel generation methods for community use.

*5.4.2 Leveraging Communication Boards for Model Training.* Another promising research direction is investigating how to leverage communication boards created by professionals to train or fine-tune computer vision models and vocabulary generation algorithms. Again, collecting data from real-world usage through the public release of AAC applications provides an opportunity to improve the quality of generated vocabulary and enhance the system's performance.

*5.4.3 Assessment of Impact on AAC Users and Professionals.* Beyond system development, further research efforts should assess the impact of AAC apps like QuickPic on various stakeholders, including SLPs, special education teachers, family members, and individuals with complex communication needs.

For SLPs, a natural progression of research would involve evaluating the time required to create communication boards using automated apps as compared to traditional tools that rely on manual programming. This assessment could encompass controlled experiments similar to the one in this paper, as well as real therapy sessions with clients. For special education settings, an intriguing avenue of research involves investigating the impact of automated AAC tools in providing support to multiple individuals simultaneously, as well as their potential to encourage peer communication. Considering the significant issue of device abandonment due to insufficient family participation in offering appropriate language

[7]https://openai.com/research/gpt-4
[8]https://ai.google/discover/palm2/
[9]https://ai.meta.com/llama/

support, future research could seek to uncover how automated topic-displays affect home communication dynamics and the level of involvement of family members in the AAC mentoring process. Finally, for individuals reliant on AAC, such as those with autism, researchers should delve into the impact of AAC apps like QuickPic on therapy outcomes. This could include assessing improvements in communication abilities and the retention of vocabulary usage skills, offering valuable insights into the application's effectiveness in clinical practice.

*5.4.4 Risks Associated with LLMs.* While the use of LLMs can pose substantial harms such as stereotyping and denigration, especially when used in decision-making processes [5, 7], we judge that the risks are relatively low in the context of automatic generation of communication boards. Given that topic-specific boards such as the ones produces by QuickPic are a complementary tool for symbolic language learning, and do not aim to replace existing *robust* communication devices designed for autonomous communication, SLPs, teachers, or family members are actively involved in board creation process. Therefore, they can analyze and vet generated vocabulary. Nevertheless, future research should investigate methods to ensure that generated vocabulary does not contain age-sensitive or potentially traumatic words for broader use cases of automated AAC.

## 6 CONCLUSION

In this work, we address the pressing challenges faced by Speech-Language Pathologists (SLPs) and special education teachers in providing effective communication support for individuals with complex communication needs, particularly autistic children. Augmentative and Alternative Communication (AAC) applications are a valuable tool in this regard, but the manual process of selecting and programming relevant, personalized vocabulary has been a time-consuming hurdle for professionals.

We introduce QuickPic, the first AAC tool co-designed with experienced SLPs and special education professionals able to automatically generate topic-specific communication boards from photographs. QuickPic's design features, rooted in user requirements, facilitate "just-in-time" language support, potentially enhancing the immersive communication experience for learners. Our study also highlights the superior appropriateness of GPT-based generation, compared to a previous method for symbolic vocabulary generation for AAC. While the previous method can be effective for certain contexts, it struggled to provide relevant words for topics commonly discussed during therapy and special education.

Our research underscores the importance of co-designing AAC tools with end-users, shedding light on unique design requirements and the need for vocabulary that aligns closely with therapy and special education topics. The success of our co-design process, backed by the enthusiastic commitment of professionals and the inclusion of various perspectives, resulted in a refined prototype ready to meet the needs of SLPs and special education teachers.

## 7 ACKLOWDEGMENTS

We would like to thank all participants, who generously shared their time and expertise in the development of this work. This research was funded by the Fonds de Recherche du Québec - Nature et Technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2018-06130], the Canada Research Chairs Program (CRC), and by AGE-WELL NCE, Canada's technology and aging network.

## REFERENCES

[1] Nicoletta Adamo-Villani, Edward Carpenter, and Laura Arns. 2006. An immersive virtual environment for learning sign language mathematics. In *ACM SIGGRAPH 2006 educators program*. 20–es.

[2] Monika Althaus, Leo MJ De Sonneville, Ruud B Minderaa, Laura GN Hensen, and Riet B Til. 1996. Information processing and aspects of visual attention in children with the DSM-III-R diagnosis "pervasive developmental disorder not otherwise specified"(PDDNOS): I. Focused and divided attention. *Child Neuropsychology* 2, 1 (1996), 17–29.

[3] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is" critical" about critical design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3297–3306.

[4] Susan Baxter, Pam Enderby, Philippa Evans, and Simon Judge. 2012. Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: a systematic review and qualitative synthesis. *International Journal of Language & Communication Disorders* 47, 2 (2012), 115–129.

[5] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[6] David Beukelman, Jackie McGinnis, and Deanna Morrow. 1991. Vocabulary selection in augmentative and alternative communication. *Augmentative and alternative communication* 7, 3 (1991), 171–185.

[7] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[8] Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Ajit Narayanan, Meredith Ringel Morris, and Michael P Brenner. 2022. Context-Aware Abbreviation Expansion Using Large Language Models. *arXiv preprint arXiv:2205.03767* (2022).

[9] Ann Copestake. 1997. Augmented and alternative NLP techniques for augmentative and alternative communication. In *Natural Language Processing for Communication Aids*.

[10] Jim Cummins. 2000. Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. (2000).

[11] Mauricio Fontana de Vargas and Karyn Moffatt. 2021. Automated Generation of Storytelling Vocabulary from Photographs for use in AAC. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 1353–1364.

[12] Patrick W Demasco and Kathleen F McCoy. 1992. Generating text from compressed input: An intelligent interface for people with severe motor impairments. *Commun. ACM* 35, 5 (1992), 68–78.

[13] Cynthia Donato, Howard C Shane, and Bronwyn Hemsley. 2014. Exploring the feasibility of the visual language in autism program for children in an early intervention group setting: Views of parents, educators, and health professionals. *Developmental neurorehabilitation* 17, 2 (2014), 115–124.

[14] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, et al. 2015. From captions to visual concepts and back. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1473–1482.

[15] Mauricio Fontana de Vargas, Jiamin Dai, and Karyn Moffatt. 2022. AAC with Automated Vocabulary from Photographs: Insights from School and Speech-Language Therapy Settings. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.

[16] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 937–946.

[17] D Jeffery Higginbotham, Howard Shane, Susanne Russell, and Kevin Caves. 2007. Access to AAC: Present, past, and future. *Augmentative and alternative communication* 23, 3 (2007), 243–257.

[18] Ting-Hao Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, et al. 2016. Visual storytelling. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1233–1239.

[19] Anna Jonsson, Lina Kristoffersson, Ulrika Ferm, and Gunilla Thunberg. 2011. The ComAlong communication boards: parents' use and experiences of aided language stimulation. *Augmentative and Alternative Communication* 27, 2 (2011), 103–116.

[20] Shaun K Kane, Barbara Linam-Church, Kyle Althoff, and Denise McCall. 2012. What we talk about: Designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 49–56.

[21] Per Ola Kristensson, James Lilley, Rolf Black, and Annalu Waller. 2020. A Design Engineering Approach for Quantitatively Exploring Context-Aware Sentence Retrieval for Nonspeaking Individuals with Motor Disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[22] Per Ola Kristensson and Thomas Müllners. 2021. Design and analysis of intelligent text entry systems with function structure models and envelope analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[23] Joanne Lasker and David R Beukelmanoe. 1999. Peers' perceptions of storytelling by an adult with aphasia. *Aphasiology* 13, 9-11 (1999), 857–869.

[24] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597* (2023).

[25] Janice Light and Kathryn Drager. 2007. AAC technologies for young children with complex communication needs: State of the science and future research directions. *Augmentative and alternative communication* 23, 3 (2007), 204–216.

[26] Janice Light and David McNaughton. 2012. Supporting the communication, language, and literacy development of children with complex communication needs: State of the science and future research priorities. *Assistive Technology* 24, 1 (2012), 34–44.

[27] Janice Light, Krista M Wilkinson, Amber Thiessen, David R Beukelman, and Susan Koch Fager. 2019. Designing effective AAC displays for individuals with developmental or acquired disabilities: State of the science and future research directions. *Augmentative and Alternative Communication* 35, 1 (2019), 42–55.

[28] Stephen Lindsay, Katie Brittain, Daniel Jackson, Cassim Ladha, Karim Ladha, and Patrick Olivier. 2012. Empathy, participatory design and people with dementia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 521–530.

[29] Stephen Lindsay, Daniel Jackson, Guy Schofield, and Patrick Olivier. 2012. Engaging older people using participatory design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1199–1208.

[30] Kathleen F McCoy, Christopher A Pennington, and Arlene Luberoff Badman. 1998. Compansion: From research prototype to practical integration. *Natural Language Engineering* 4, 1 (1998), 73–95.

[31] Aimee Mooney, Steven Bedrick, Glory Noethe, Scott Spaulding, and Melanie Fried-Oken. 2018. Mobile technology to support lexical retrieval during activity retell in primary progressive aphasia. *Aphasiology* 32, 6 (2018), 666–692.

[32] Alison Moorcroft, Nerina Scarinci, and Carly Meyer. 2019. Speech pathologist perspectives on the acceptance versus rejection or abandonment of AAC systems for children with complex communication needs. *Augmentative and Alternative Communication* 35, 3 (2019), 193–204.

[33] Norashikin Mustafa, Nik Shanita Safii, Aida Jaffar, Nor Samsiah Sani, Mohd Izham Mohamad, Abdul Hadi Abd Rahman, Sherina Mohd Sidik, et al. 2021. Malay version of the mHealth app usability questionnaire (M-MAUQ): translation, adaptation, and validation study. *JMIR mHealth and uHealth* 9, 2 (2021), e24457.

[34] Timothy Neate, Aikaterini Bourazeri, Abi Roper, Simone Stumpf, and Stephanie Wilson. 2019. Co-created personas: Engaging and empowering users with diverse needs within the design process. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[35] Timothy Neate, Abi Roper, Stephanie Wilson, Jane Marshall, and Madeline Cruice. 2020. CreaTable content and tangible interaction in Aphasia. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[36] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. 2021. Designing AACs for People with Aphasia Dining in Restaurants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[37] Amanda O'Brien, Meghan O'Brien, Ralf W Schlosser, Christina Yu, Anna A Allen, Suzanne Flynn, John Costello, and Howard C Shane. 2017. Repurposing consumer products as a gateway to just-in-time communication. In *Seminars in Speech and Language*, Vol. 38. Thieme Medical Publishers, 297–312.

[38] Ehud Reiter, Ross Turner, Norman Alm, Rolf Black, Martin Dempster, and Annalu Waller. 2009. Using NLG to help language-impaired users tell stories and participate in social dialogues. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*. 1–8.

[39] Michael Scaife, Yvonne Rogers, Frances Aldrich, and Matt Davies. 1997. Designing for or designing with? Informant design for interactive learning environments. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 343–350.

[40] Ralf W Schlosser, Howard C Shane, Anna A Allen, Jennifer Abramson, Emily Laubscher, and Katherine Dimery. 2016. Just-in-time supports in augmentative and alternative communication. *Journal of Developmental and Physical Disabilities* 28, 1 (2016), 177–193.

[41] Ralf W Schlosser, Howard C Shane, Anna A Allen, Andrea Benz, Jacqueline Cullen, Lindsay O'Neill, Laurel Chiesa, Lisa Miori-Dinneen, Rajinder Koul, and Rubini Pasupathy. 2020. Coaching a school team to implement the Visual Immersion System™ in a classroom for children with autism spectrum disorder: A mixed-methods proof-of-concept study. *Advances in Neurodevelopmental Disorders* 4 (2020), 447–470.

[42] Howard C Shane and Patti Ducoff Albert. 2008. Electronic screen media for persons with autism spectrum disorders: Results of a survey. *Journal of autism and developmental disorders* 38 (2008), 1499–1508.

[43] Howard C Shane, Emily Laubscher, Ralf W Schlosser, Holly L Fadie, James F Sorce, Jennifer S Abramson, Suzaane Flynn, and Kara Corley. 2014. *Enhancing communication for individuals with autism: A guide to the visual immersion system*. Brookes.

[44] Howard C Shane, Emily H Laubscher, Ralf W Schlosser, Suzanne Flynn, James F Sorce, and Jennifer Abramson. 2012. Applying technology to visually support language and communication in individuals with autism spectrum disorders. *Journal of autism and developmental disorders* 42 (2012), 1228–1235.

[45] Junxiao Shen, Boyin Yang, John J Dudley, and Per Ola Kristensson. 2022. Kwickchat: A multi-turn dialogue system for aac using context-aware sentence generation by bag-of-keywords. In *27th International Conference on Intelligent User Interfaces*. 853–867.

[46] Donghoon Shin, Jaeyoon Song, Seokwoo Song, Jisoo Park, Joonhwan Lee, and Soojin Jun. 2020. TalkingBoogie: collaborative mobile AAC system for non-verbal children with developmental disabilities and their caregivers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[47] Amy Starble, Tiffany Hutchins, Mary Alice Favro, Patricia Prelock, and Brooke Bitner. 2005. Family-centered intervention and satisfaction with AAC device training. *Communication Disorders Quarterly* 27, 1 (2005), 47–54.

[48] Nava Tintarev, Ehud Reiter, Rolf Black, Annalu Waller, and Joe Reddington. 2016. Personal storytelling: Using Natural Language Generation for children with complex communication needs, in the wild.... *International Journal of Human-Computer Studies* 92 (2016), 1–16.

[49] Keith Trnka, John McCaw, Debra Yarrington, Kathleen F McCoy, and Christopher Pennington. 2009. User interaction with word prediction: The effects of prediction quality. *ACM Transactions on Accessible Computing (TACCESS)* 1, 3 (2009), 1–34.

[50] Keith Trnka, Debra Yarrington, John McCaw, Kathleen F McCoy, and Christopher Pennington. 2007. The effects of word prediction on communication rate for AAC. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*. 173–176.

[51] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K Kane. 2023. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[52] Gail M Van Tatenhove. 2009. Building language competence with students using AAC devices: Six challenges. *Perspectives on Augmentative and Alternative Communication* 18, 2 (2009), 38–47.

[53] Keith Vertanen. 2017. Towards improving predictive aac using crowdsourced dialogues and partner context. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 347–348.

[54] Cara Wilson, Margot Brereton, Bernd Ploderer, and Laurianne Sitbon. 2019. Co-Design Beyond Words: 'Moments of Interaction'with Minimally-Verbal Children on the Autism Spectrum. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[55] Cara Wilson, Laurianne Sitbon, Bernd Ploderer, Jeremy Opie, and Margot Brereton. 2020. Self-expression by design: Co-designing the expressiball with minimally-verbal children on the autism spectrum. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[56] Stephanie Wilson, Abi Roper, Jane Marshall, Julia Galliers, Niamh Devane, Tracey Booth, and Celia Woolf. 2015. Codesign for people with aphasia through tangible design languages. *CoDesign* 11, 1 (2015), 21–34.

[57] Leming Zhou, Jie Bao, I Made Agus Setiawan, Andi Saptono, Bambang Parmanto, et al. 2019. The mHealth App Usability Questionnaire (MAUQ): development and validation study. *JMIR mHealth and uHealth* 7, 4 (2019), e11500.

[58] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502.